Delgado, M. A., Escanciano, J. C. (2016). Distribution-free tests of conditional moment inequalities. Journal of Statistical Planning and Inference, v. 173, pp. 99-108.
Available in: https://doi.org/10.1016/j.jspi.2015.12.005

Distribution-free tests of conditional moment inequalities☆

Miguel A. Delgado[a,*], Juan Carlos Escanciano[b]

[a] Departamento de Economía, Universidad Carlos III de Madrid, 28903 Getafe (Madrid), Spain

[b] Economics Department, Indiana University, Bloomington, IN, USA

A B S T R A C T

This article proposes testing the hypothesis of a uniformly non-positive nonparametric regression function using a test statistic with tabulated critical values. The null hypothesis is characterized in terms of the significance of a parameter, which measures a distance from the double-integrated regression function to the class of concave functions. The test statistic is a suitably scaled parameter estimate, which does not require smooth estimation of the underlying regression and/or the conditional variance functions. The finite sample performance of the proposed test is studied by means of two Monte Carlo experiments, showing that the proposed method compares favorably to existing procedures.

Keywords:
Nonparametric testing
Conditional inequalities
Least concave majorant

1. Introduction and summary

Let (Y, X) be a bivariate random vector defined on $(\Omega, \mathcal{A}, \mathbb{P})$. Assume Y is integrable so that the regression function $m(X) := \mathbb{E}\,(Y|X)$ is well defined almost surely (*a.s.*). This article proposes testing the hypothesis

$$H_0 : m(X) \leq 0 \quad a.s., \tag{1}$$

in the direction of non-parametric alternatives H_1, which consists of all cases where H_0 is not satisfied.

Inequality restrictions such as (1) appear naturally when testing treatment effects controlling for covariates. Let D be an indicator of participation in a treatment program, i.e. $D = 1$ if the individual participates in the program and 0 otherwise. Denote the observed outcome by $Z = Z(1)D + Z(0)(1-D)$, where $Z(1)$ and $Z(0)$ are the potential outcomes with and without treatment, respectively. The treatment is successful uniformly in the covariate X, e.g. age, if $\mathbb{E}\,(Z(0) - Z(1)|X) \leq 0$ *a.s.*, which can be expressed as (1) with $Y = (\mathbb{E}\,(D|X) - D)\,Z$, provided $0 < \mathbb{E}\,(D|X) < 1$ *a.s.* and the treatment is randomized conditional on covariates, i.e. $Z(1)$ and $Z(0)$ are independent of D, conditional on X. See Delgado and Escanciano (2013) and Chang et al. (2015) for further discussion. Identifiability conditions on econometric models often appear as testable restrictions on moment inequalities. For instance, behavioral choice models generate conditional moment inequalities suitable to identifying parameters of nonparametric functions of interests; see Pakes (2010). This includes testing the “realistic expectation hypothesis” in insurance market modeling; e.g. Chiappori et al. (2006). Inference on game theoretical models often assume that some underlying regression function is non-negative; see de Paula (2013) for a survey. Inequality restrictions on conditional models also arise when testing revealed preferences; see e.g. Blundell et al. (2003). Finally, partial

☆ We are most thankful to the two referees, whose many insightful comments and suggestions have led to a much improved version of the article. Research funded by Spanish Ministerio de Economía y Competitividad; Plan Nacional de I+D+i, reference number ECO2014-55858-P, and Unidad de Excelencia María de Maeztu reference number MDM-2014-0431.

* Corresponding author.
E-mail addresses: miguelangel.delgado@uc3m.es (M.A. Delgado), jescanci@indiana.edu (J.C. Escanciano).

identification conditions can often be written as conditional moment inequalities. Some references on inference procedures on moment inequalities are Khan and Tamer (2009), Chernozhukov et al. (2013), Armstrong (2015) and references therein.

Under (1), $\mathbb{E}\left(Y \cdot 1_{\{a \le X \le b\}}\right) \le 0$ for all $a, b \in \mathbb{R}$. This fact has suggested tests of (1) based on local averages. Dümbgen and Spokoiny (2001) and Juditsky and Nemirovski (2002) proposed a test of qualitative hypotheses on the signal of a Gaussian white noise model, which include positivity, based on kernel estimators of the regression function. The resulting test is adaptive in the class of smooth functions considered. Baraud et al. (2003, 2005) proposed a test based on trimmed averages for qualitative hypotheses on the regression function of a fixed design model with homoskedastic errors. Exact critical values of these tests are derived under Gaussian errors. Asymptotic tests of positivity in the context of general models with random covariates and possibly non Gaussian errors have been recently proposed by Kim (2008), Andrews and Shi (2013), Chetverikov (2013) and Armstrong (2015), among others. The critical values of these tests must be estimated with the assistance of bootstrap techniques. The test of Lee et al. (2013) is of a different nature. The test statistic is based on a one-sided version of the L^p*-type* functionals of kernel estimators using standard normal critical values. The asymptotic test is justified when the bandwidth converges to zero at a suitable rate related to the sample size and assuming different restrictions on m. See also Lee et al. (2014). The bootstrap test of Delgado and Escanciano (2013), related to Durot (2003) and Delgado and Escanciano (2011) monotonicity tests, avoids estimating the regression function. In the general case, the limiting distribution of their test depends on a nuisance parameter, the integrated conditional variance. This article applies this testing methodology to construct a test with pivotal critical values, free of nuisance and tuning parameters.

Henceforth, let F denote the cumulative distribution function (cdf) of X, which is assumed to be continuous, and for a generic monotone function $G : \mathbb{R} \to \mathbb{R}$, let G^{-1} denote its generalized inverse $G^{-1}(r) := \inf\{t \in \mathbb{R} : G(t) \ge r\},\ r \in \mathbb{R}$. The null hypothesis can be equivalently expressed as

$$H_0 : M \text{ is non-increasing},$$

where

$$M(u) = \mathbb{E}\left[Y \cdot 1_{\{F(X) \le u\}}\right] = \int_0^u \left(m \circ F^{-1}\right)(v)dv, \quad u \in [0, 1]$$

is the integrated regression function, and $\circ$ denotes composition of functions. This, in turn, is satisfied if

$$H_0 : \mathbb{M} \text{ is concave},$$

where

$$\mathbb{M}(u) := \int_0^u M(v)dv, \quad u \in [0, 1]\,.$$

We exploit this fact, expressing H_0 as a significance test on a parameter by using the least concave majorant (lcm) operator $\mathcal{L}$, defined as follows. For any function $g : [0, 1] \to \mathbb{R}$, (i) $\mathcal{L}g$ is concave and (ii) if there exists a concave function h with $h \ge g$, then $h \ge \mathcal{L}g$. Let $\|\cdot\|$ be a norm defined on the space of continuous functions satisfying the Riesz's property, i.e. if $0 \le g(u) \le h(u)$, for all $u \in [0, 1]$, then $\|g\| \le \|h\|$. Examples of possible norms $\|\cdot\|$ include the sup-norm $\|g\|_\infty := \sup_{u \in [0,1]} |g(u)|$ and the L_2-norm $\|g\|_2^2 := \int_0^1 g^2(u)du$. The hypotheses can be alternatively expressed in terms of the parameter $\eta = \|\mathcal{L}\mathbb{M} - \mathbb{M}\| \ge 0$, i.e.

$$H_0 : \eta = 0 \quad \text{vs.} \quad H_1 : \eta > 0.$$

The parameter η measures a distance from m to the class of non-negative functions.

Given a random sample $\{(Y_i, X_i)\}_{i=1}^n$ of independent and identically distributed (i.i.d.) copies of (Y, X), the test statistic is based on an estimator of η. First, the integrated regression function $M(u)$ is estimated by,

$$\hat{M}(u) = \frac{1}{n}\sum_{i=1}^n Y_i \cdot 1_{\left\{\hat{F}(X_i) \le u\right\}},$$

where $\hat{F}(\cdot) := n^{-1}\sum_{i=1}^n 1_{\{X_i \le \cdot\}}$ is the empirical analog of F. Henceforth, we do not indicate the dependence of the statistics on the sample size n. This suggests the use of the following estimator of η

$$\hat{\eta} = \left\|\mathcal{L}\widehat{\mathbb{M}} - \widehat{\mathbb{M}}\right\|,$$

with

$$\widehat{\mathbb{M}}(u) = \int_0^u \hat{M}(v)dv = \frac{1}{n}\sum_{i=1}^n Y_i \cdot \left[u - \hat{F}(X_i)\right] 1_{\left\{\hat{F}(X_i) \le u\right\}}.$$

Since $\hat{\eta}$ is expected to take small values under H_0 and large values under H_1, a scaled version of $\hat{\eta}$ could be used as a test statistic. A related testing strategy was suggested by Durot (2003) in order to test that m is monotonic in the context of a fixed design model with homoskedastic errors. In the general case, asymptotic critical values for tests based on $\hat{\eta}$ depend on the integrated variance $\tau(u) := \int_0^u \left(\sigma^2 \circ F^{-1}\right)(v)dv$, where $\sigma^2(\cdot) := \operatorname{Var}(Y|X = \cdot)$. Delgado and Escanciano (2013) suggested a bootstrap test using $\hat{\eta}$ as test statistic. In contrast, this article proposes a modification of $\hat{\eta}$, so that the resulting test is asymptotically pivotal. The main contributions of the article are summarized as follows:

(i) Derivation of suitable critical values for tests of moment inequalities based on the lcm;
(ii) Construction of a new estimator for the integrated conditional variance τ based on partial sums of squared successive differences of concomitants (see von Neumann et al., 1941);
(iii) Construction of a modified test with asymptotically distribution-free critical values;
(iv) Derivation of the limiting distribution of the proposed test under the least favorable case (lfc) in the null hypothesis and computation of some of its quantiles;
(v) Proof of the consistency of the test; and
(vi) Simulations showing a satisfactory finite sample performance of the proposed test, comparing favorably to existing procedures.

The rest of the article is organized as follows. Next section deals with (i), while Section 3 addresses (ii–v). Section 4 discusses implementation of the test with standard software routines and reports the results of Monte Carlo experiments (vi). Suggestions for further research can be found in Section 5. Proofs are gathered in an Appendix.

2. Justification of critical values

In this section we justify the use of certain critical values for the test statistic $\hat{\eta}$ and related statistics. The asymptotic distribution of $\hat{\eta}-\eta$ at any circumstance under H_0 is hard to derive (see Beare and Moon, 2015, for a related problem). In contrast, the asymptotic distribution of

$$\bar{\eta}=\left\|\mathcal{L}\left(\widehat{\mathbb{M}}-\widehat{\mathbb{M}}_0\right)-\left(\widehat{\mathbb{M}}-\widehat{\mathbb{M}}_0\right)\right\|,$$

with

$$\widehat{\mathbb{M}}_0(u):=\int_0^u \hat{M}_0(v)dv$$

and

$$\hat{M}_0(u):=\frac{1}{n}\sum_{i=1}^{n} m(X_i)1_{\left\{\hat{F}(X_i)\le u\right\}},$$

is easy to obtain by applying the invariance principle for concomitants. Notice that,

$$\begin{aligned}\bar{T}(u)&=\left(\hat{M}-\hat{M}_0\right)(u)\\ &=\frac{1}{n}\sum_{i=1}^{\lfloor nu\rfloor}\varepsilon_{(i:n)},\end{aligned}$$

where, henceforth, $\varepsilon_i = Y_i - m(X_i)$ are the innovations, and for any set of generic random variables $\{\xi_i\}_{i=1}^n$, $\xi_{(i:n)}$ is the concomitant of the *i-th* order statistic $X_{i:n}$, i.e. $\xi_{(i:n)}=\xi_j$ iff $X_{i:n}=X_j$ with $X_{1:n}<X_{2:n}<\cdots<X_{n:n}$. Since F is continuous, the strict inequalities in the order statistics occur with probability one. The empirical process $\bar{T}$ is the cumulative sum (CUSUM) of ε-concomitants to the order statistics of the covariates. Bhattacharya (1974, Lemma 1), shows that $\varepsilon_{(1:n)},\ldots,\varepsilon_{(n:n)}$ are conditionally independent given $\{X_i\}_{i=1}^n$, with conditional distribution $G(\cdot|X_{1:n}),\ldots,\ G(\cdot|X_{n:n})$, respectively, where $G(\cdot|x)=\mathbb{P}(\varepsilon_1\le\cdot|X_1=x)$. Using this result, Bhattacharya (1974, Theorem 2, 1984, Lemma 9.2) shows that, under suitable regularity conditions to be discussed in the next section,

$$\sqrt{n}\bar{T}\to_d B\circ\tau, \tag{2}$$

where B is a standard Brownian motion on $[0,1]$, and, henceforth, $\to_d$ denotes convergence in distribution in the Skorohod's space $D[0,1]$. Therefore, by the continuous mapping theorem (cmt),

$$\sqrt{n}\left(\widehat{\mathbb{M}}-\widehat{\mathbb{M}}_0\right)\to_d \mathbb{B}_\tau$$

with $\mathbb{B}_\tau(u)\overset{d}{=}\int_0^u(B\circ\tau)(v)dv$ ($\overset{d}{=}$ denotes equality in distribution), and

$$\sqrt{n}\bar{\eta}\to_d\zeta_\tau\overset{d}{=}\|\mathcal{L}\mathbb{B}_\tau-\mathbb{B}_\tau\|.$$

Next Proposition, justifies the asymptotic validity of the α significance level test $\hat{\Phi}(c_{\tau\alpha})=1_{\{\sqrt{n}\cdot\hat{\eta}>c_{\tau\alpha}\}}$ with $\mathbb{P}(\zeta_\tau>c_{\tau\alpha})=\alpha$ (note that the distribution of ζ_τ is continuous, see Lifshits, 1982). Next result uses the fact that $\hat{M}_0$ is non-increasing under H_0, which implies that $\widehat{\mathbb{M}}_0$ is concave under H_0.

Proposition 1. *Under H_0, $\bar{\eta}\ge\hat{\eta}$ a.s. and $\bar{\eta}=\hat{\eta}$ a.s. if $m(X)=0$ a.s.*

The proposition, as well as the rest of results in the article are proved in the Appendix. Consider the power function $\hat{\beta}_\alpha = \lim_{n\to\infty} \mathbb{E}\left[\hat{\Phi}(c_{\tau\alpha})\right]$. Proposition 1 guarantees that $\hat{\beta}_\alpha \le \alpha$ under H_0, and $\hat{\beta}_\alpha = \alpha$ when $m(X) = 0$ *a.s.* Under H_1, $\hat{\beta}_\alpha = 1$, since $\hat{\eta} \to_p \eta > 0$. This justifies the asymptotic validity of the test. However, $c_{\tau\alpha}$ is case dependent, since it depends on the unknown function τ, and must be estimated.

Under homoskedasticity, i.e. when $\sigma^2(X) = \mathbb{E}\left(\varepsilon_1^2\right) = \tau(1)$ *a.s.*,

$$\sqrt{n}\frac{\bar{\eta}}{\sqrt{\tau(1)}} \to_d \zeta_0 \overset{d}{=} \|\mathcal{L}\mathbb{B} - \mathbb{B}\|,$$

where $\mathbb{B}(u) \overset{d}{=} \int_0^u B(v)dv$, since $\tau(u) = u \cdot \tau(1)$. Therefore, the distribution of ζ_0 is pivotal, which suggests a test using $\sqrt{n}\hat{\eta}\big/\sqrt{\tau(1)}$ as test statistic and critical values based on the quantiles of ζ_0. The error variance $\tau(1) = \mathbb{E}\left(\varepsilon_1^2\right)$ is unknown, but it can be estimated using the sample mean of squared successive differences of concomitants,

$$\tilde{\tau}(1) := \frac{1}{2(n-1)}\sum_{i=2}^{n}\left(Y_{(i:n)} - Y_{(i-1:n)}\right)^2. \tag{3}$$

See von Neumann et al. (1941) in the context of nonparametric regression with a fixed design (see also Rice, 1984; Gasser et al., 1986 or Hall et al., 1990 for further developments). In the next section, we show that this estimator is consistent under suitable regularity conditions. Therefore, under homoskedasticity, the test statistic $\sqrt{n}\hat{\eta}\big/\sqrt{\tilde{\tau}(1)}$ with the critical values of ζ_0, which can be tabulated, provide a valid test. This test can be applied to testing positivity of the regression under a fixed design with homoskedastic disturbances, like Durot's (2003) monotonicity test.

In the next section, we propose an estimator of τ, inspired by (3), which is used to transform $\bar{T}$ into an asymptotically distribution-free empirical process, which forms a basis to construct an asymptotically pivotal test in the presence of heteroskedasticity, using an alternative test statistic to $\hat{\eta}$ whose critical values are those of ζ_0.

3. Asymptotically distribution-free tests

Notice that, by the scale invariance property of Brownian motion,

$$(B \circ \tau)(u) \overset{d}{=} \sqrt{\tau(1)} \cdot (B \circ \kappa)(u),$$

where, under our conditions below, $\kappa(u) := \tau(u)\big/\tau(1)$ is a monotonically increasing and continuous function on $[0, 1]$. Therefore,

$$\frac{\sqrt{n}}{\sqrt{\tau(1)}}\left(\bar{T} \circ \kappa^{-1}\right) \to_d B, \tag{4}$$

which suggests characterizing H_0 by means of the parameter $\eta_\tau = \|\mathcal{L}\mathbb{M}_\tau - \mathbb{M}_\tau\|$, with $\mathbb{M}_\tau(u) = \int_0^u M_\tau(v)dv$ and

$$M_\tau(u) = \frac{\left(M \circ \kappa^{-1}\right)(u)}{\sqrt{\tau(1)}} = \frac{\mathbb{E}\left(Y \cdot 1_{\{(\kappa \circ F(X)) \le u\}}\right)}{\sqrt{\tau(1)}}.$$

So, H_0 can be alternatively characterized as

$$H_0 : \eta_\tau = 0 \quad \text{versus} \quad H_1 : \eta_\tau > 0.$$

A natural estimator of $\tau(u)$, for $u \in [0, 1]$, inspired by (3), is

$$\tilde{\tau}(u) := 1_{\{u\in[2/n,1]\}}\frac{1}{2(n-1)}\sum_{i=2}^{\lfloor nu \rfloor}\left(Y_{(i:n)} - Y_{(i-1:n)}\right)^2.$$

The integrated conditional variance function τ appears in other inference problems such as in the context of model checking, see e.g. Koul and Stute (1999), who considered a parametric estimator of τ based on regression residuals. To the best of our knowledge neither $\tilde{\tau}$, or any other estimator of τ with no preliminary estimation of m, has been considered before in the literature. We first show that $\tilde{\tau}$ is a consistent estimator for τ, uniformly on $u \in [0, 1]$, under the following condition.

A1. (i) F is continuous; (ii) $\gamma(\cdot) = \mathbb{E}\left((Y - m(X))^4 \,|X = \cdot\right)$ is uniformly bounded; and (iii) $\sigma^2(\cdot)$ is of bounded variation and bounded away from zero.

A2. $m \circ F^{-1}$ is continuous on $[0, 1]$.

Theorem 1. *Under* A1–A2, $\sup_{u\in[0,1]} |\tilde{\tau}(u) - \tau(u)| = o_{\mathbb{P}}(1)$.

The parameter $\eta_\tau = \|\mathcal{L}\mathbb{M}_\tau - \mathbb{M}_\tau\|$ is estimated by

$$\tilde{\eta}_\tau = \left\|\mathcal{L}\widetilde{\mathbb{M}}_\tau - \widetilde{\mathbb{M}}_\tau\right\|$$

with $\widetilde{\mathbb{M}}_\tau(u) = \int_0^u \tilde{M}_\tau(v)dv$ and

$$\tilde{M}_\tau(u) = \frac{\left(\hat{M} \circ \tilde{\kappa}^{-1}\right)(u)}{\sqrt{\tilde{\tau}(1)}} = \frac{1}{n\sqrt{\tilde{\tau}(1)}} \sum_{i=1}^{n} Y_i \cdot 1_{\left\{\left(\tilde{\kappa}\circ\hat{F}\right)(X_i)\le u\right\}},$$

where $\tilde{\kappa}(u) = \tilde{\tau}(u)\big/\tilde{\tau}(1)$. It is computationally worth noticing that

$$\widetilde{\mathbb{M}}_\tau(u) = \frac{1}{n\sqrt{\tilde{\tau}(1)}} \sum_{i=1}^{n} Y_i \cdot \left[u - \left(\tilde{\kappa} \circ \hat{F}\right)(X_i)\right] \cdot 1_{\left\{\left(\tilde{\kappa}\circ\hat{F}\right)(X_i)\le u\right\}}.$$

Likewise, define

$$\bar{\eta}_\tau = \left\|\mathcal{L}\left(\widetilde{\mathbb{M}}_\tau - \widetilde{\mathbb{M}}_{0\tau}\right) - \left(\widetilde{\mathbb{M}}_\tau - \widetilde{\mathbb{M}}_{0\tau}\right)\right\|,$$

with $\widetilde{\mathbb{M}}_{0\tau}(u) = \int_0^u \tilde{M}_{0\tau}(v)dv$ and

$$\tilde{M}_{0\tau}(u) = \frac{\left(\hat{M}_0 \circ \tilde{\kappa}^{-1}\right)(u)}{\sqrt{\tilde{\tau}(1)}} = \frac{1}{n\sqrt{\tilde{\tau}(1)}} \sum_{i=1}^{n} m(X_i) \cdot 1_{\left\{\left(\tilde{\kappa}\circ\hat{F}\right)(X_i)\le u\right\}}.$$

Next Theorem justifies the test $\tilde{\Phi}(c_{0\alpha}) = 1_{\{\sqrt{n}\cdot\tilde{\eta}_\tau > c_{0\alpha}\}}$ where $c_{0\alpha}$ is the $(1-\alpha) - th$ quantile of ζ_0. The theorem is proved in the Appendix by showing that $\sqrt{n}\bar{\eta}_\tau \to_d \zeta_0,\ \tilde{\eta}_\tau \to_p \eta_\tau$ and $\bar{\eta}_\tau \ge \tilde{\eta}_\tau$ *a.s.* under H_0, with strict equality when $m(X) = 0$ *a.s.* Define the asymptotic power function $\tilde{\beta}_\alpha := \lim_{n\to\infty} \mathbb{E}\left[\tilde{\Phi}(c_{0\alpha})\right]$.

Theorem 2. *Under* A1–A2, $\tilde{\beta}_\alpha \le \alpha$ *under* H_0, $\tilde{\beta}_\alpha = \alpha$ *when* $m(X) = 0$ *a.s., and* $\tilde{\beta}_\alpha = 1$ *under* H_1.

Next Section discusses the practical implementation of the test and its finite sample performance.

4. Computational issues and finite sample properties

This section discusses computational aspects and reports the results of some Monte Carlo experiments to investigate the finite sample performance of the proposed method. Our theoretical results are valid for any norm $\|\cdot\|$ satisfying the Riesz property, but one norm that leads to computationally simple calculations is the sup-norm $\|\cdot\|_\infty$. Thus, in the rest of this section we focus on this norm.

First, note that $\widetilde{\mathbb{M}}_\tau$ is piecewise linear with jumps at the points $\tilde{\kappa}_i \equiv \tilde{\kappa}(i/n)$, for $i = 1, \ldots, n$, and therefore, we can express

$$\widetilde{\mathbb{M}}_\tau(\tilde{\kappa}_\ell) = \sum_{j=1}^{l} \tilde{r}_j \tilde{w}_j, \quad l = 1, \ldots, n,$$

where $\tilde{r}_1 \equiv 0,\ \tilde{w}_1 \equiv \tilde{\kappa}_1$ and for $j \ge 2$, $\tilde{w}_j = (\tilde{\kappa}_j - \tilde{\kappa}_{j-1})$ and

$$\tilde{r}_j \equiv \frac{1}{n\sqrt{\tilde{\tau}(1)}} \sum_{i=1}^{j-1} Y_{(i:n)}, \quad j = 2, \ldots, n. \tag{5}$$

In turn, this representation of $\widetilde{\mathbb{M}}_\tau$ is useful to compute $\mathcal{L}\widetilde{\mathbb{M}}_\tau$, as the knots of $\mathcal{L}\widetilde{\mathbb{M}}_\tau$ are easily located applying the Pooled Adjacent Violators Algorithm (PAVA) proposed by Barlow et al. (1972). The input for the algorithm is $\left\{\tilde{r}_i, \tilde{w}_i\right\}_{i=1}^n$, which can be computed recursively according to (5). See Cran (1980) and Bril et al. (1984) for FORTRAN implementations and de Leeuw et al. (2009) for R routines. Many statistical softwares have already existing code for computing cumulative sums, which makes the computation of the test statistic fast and straightforward. Consideration of the sup norm $\|\cdot\|_\infty$ leads to a Kolmogorov–Smirnov (KS) type test statistic, with

$$\tilde{\eta}_\tau = \max_{1\le i\le n} \sqrt{n}\left(\mathcal{L}\widetilde{\mathbb{M}}_\tau - \widetilde{\mathbb{M}}_\tau\right)(\tilde{\kappa}_i). \tag{6}$$

Matlab code for computing our KS test is available from the authors upon request.

Although the proposed KS test is asymptotically distribution-free, its critical values have not been tabulated previously in the literature. An analytical approach (e.g. based on the Paley–Wiener approximation, see Paley and Wiener, 1934) seems difficult in this setting, due to the nonlinearity of the lcm operator. A more practical approach can be based on an

Table 1

Rejection probabilities for test.

n	50			100			1000		
Model/α	10%	5%	1%	10%	5%	1%	10%	5%	1%
(i)	0.100	0.051	0.013	0.093	0.050	0.010	0.098	0.051	0.010
(ii)	0.108	0.058	0.015	0.102	0.054	0.013	0.100	0.052	0.013
(iii)	0.100	0.053	0.013	0.094	0.047	0.011	0.095	0.049	0.009
(iv)	0.510	0.379	0.180	0.725	0.590	0.342	1.000	1.000	1.000
(v)	0.090	0.047	0.010	0.087	0.040	0.008	0.365	0.154	0.018
(vi)	0.709	0.565	0.304	0.950	0.883	0.652	1.000	1.000	1.000

Note: 10 000 Monte Carlo simulations.

approximation of the Brownian motion by a partial sum of i.i.d. standard normals, since the lcm can be computed by the PAVA algorithm. That is, we approximate the limiting null distribution of (6) with the Monte Carlo distribution of

$$\max_{1\le i\le n} \sqrt{n}\left(\mathcal{L}W(i/n) - W(i/n)\right),$$

where

$$W(u) = \frac{1}{n}\sum_{i=1}^{\lfloor nu \rfloor} Z_i \left(u - \frac{i}{n}\right),$$

and where $\{Z_i\}_{i=1}^{n}$ are i.i.d. standard normals and n is the number of draws. The Monte Carlo approximation of the critical values for the KS test with $n = 100{,}000$ and one million simulations are respectively, 0.2051, 0.2532 and 0.3463, at 10%, 5% and 1% nominal level. We use these simulated critical values to evaluate the finite sample performance of the KS test through Monte Carlo simulations using 10,000 Monte Carlo experiments. We report rejection probabilities at 10%, 5% and 1% significance levels for different sample sizes.

We first investigate the size accuracy and power of the proposed KS tests for the following designs:

(i) $Y = u$;
(ii) $Y = cXu$;
(iii) $Y = c\exp(-0.5X)u$;
(iv) $Y = X + cXu$;
(v) $Y = 8(X - 0.5)^3 + c\exp(-0.5X)u$;
(vi) $Y = \sin(2\pi X) + c\exp(-0.5X)u$;

where X is distributed as $U[0, 1]$, independently of the standard normal error u, and c is a scale constant such that the unconditional variance of the regression error ε is $\sigma^2 = 1$. Table 1 reports the proportion of rejections for models (i)–(vi) and several sample sizes.

Models (i)–(iii) fall under the null hypothesis. Model (i) is homoskedastic, whereas models (ii) and (iii) are heteroskedastic. The KS test exhibits good size accuracy, even for small sample sizes as $n = 50$. The test is robust to the presence of heteroscedasticity. The power is moderate for $n = 50$ under the alternative (iv), is low for (v) and good for (vi). The alternative (v) is hard to detect, and requires samples sizes of at least $n = 2000$ to achieve a 0.5 rejection probability at 5%. In unreported results, we observed that for $n = 4000$ the test already rejects in 100% of the cases, which confirms its consistency.

We now compare the performance of the proposed test with the tests considered by Andrews and Shi (2013) and Lee et al. (2013) for testing conditional inequalities. We use the same designs considered in Lee et al. (2013):

DGP0: $Y = u$;
DGP1-5: $Y = X(1 - X) - c + u$;

where X is distributed as $U[0, 1]$, independently of the standard normal error u, and $c \in \{0.25, 0.20, 0.15, 0.10, 0.05\}$. We also consider heteroskedastic versions where $X \cdot u$ replaces u above (i.e. where the conditional variance is X^2). We report rejection probabilities at 5% significance level, and sample sizes $n = 50$, 200 and 1000. Thus, the designs, parameters, and sample sizes in the Monte Carlo are the same as those used in Lee et al. (2013). We refer the reader to this reference for the finite sample performance of the L_1-test of Lee et al. (2013) and the integrated based test of Andrews and Shi (2013). Table 2 reports the proportion of rejections for designs and sample sizes considered in Lee et al. (2013).

DGP0–DGP1 ($c = 0.25$) fall under the null hypothesis. Our test exhibits excellent size accuracy for DGP0, particularly for the heteroskedastic cases with $n = 200$ and $n = 1000$ (where the empirical size equals the nominal size to the third decimal). The empirical sizes of our test compare favorably with those of the test of Lee et al. (2013), which leads to some overrejections in their simulations even for large sample sizes as $n = 1000$, and they are comparable to the empirical sizes of the bootstrap test of Andrews and Shi (2013). For DGP1, corresponding to $c = 0.25$, the regression $m < 0$, and, as expected, the rejection probabilities converge to zero as n becomes large. Models DGP2-5 fall under the alternative. The empirical power is larger for lower $c's$; it is low for $c > 0.15$, moderate for $c = 0.15$ and good for $c = 0.10$ and 0.05. Our

Table 2
Rejection probabilities at 5% for test.

$\sigma^2(\cdot)$	Homoskedastic			Heteroskedastic		
Model/n	50	200	1000	50	200	1000
DGP0(lfc)	0.054	0.044	0.051	0.059	0.050	0.050
DGP1($c = 0.25$)	0.030	0.014	0.003	0.026	0.006	0.000
DGP2($c = 0.20$)	0.053	0.057	0.079	0.053	0.038	0.022
DGP3($c = 0.15$)	0.094	0.160	0.478	0.095	0.127	0.361
DGP4($c = 0.10$)	0.158	0.358	0.902	0.167	0.328	0.907
DGP5($c = 0.05$)	0.235	0.601	0.994	0.253	0.587	0.998

Note: 1000 Monte Carlo simulations.

test compares favorably with those of Lee et al. (2013) and Andrews and Shi (2013) for homoskedastic models, and it has slightly less power than Lee et al.'s (2013) test for heteroskedastic models with small values of c. Overall, these Monte Carlo simulations show that our test has an excellent performance in terms of empirical size accuracy, and it is comparable to existing tests in terms of power performance, while being simpler to implement.

5. Conclusions

We have proposed a methodology for testing inequality constraints on the regression function, which can be applied to other interesting problems. Durot (2003) proposed a test of monotonicity of m using the fact that

$$H_0 : m(X + c) - m(X) \leq 0 \quad a.s. \text{ for any } c > 0,$$

can be equivalently expressed in terms of

$$H_0 : \|\mathcal{L}M - M\|_\infty = 0.$$

Durot's test is based on the statistic $\sqrt{n}\left\|\mathcal{L}\hat{M}_\tau - \hat{M}_\tau\right\|_\infty / \sqrt{\tilde{\tau}(1)}$ and critical values based on the distribution of $\|\mathcal{L}B - B\|_\infty$. See also Delgado and Escanciano (2011) for a related approach. Our proposal suggests the tests statistic $\sqrt{n}\left\|\mathcal{L}\tilde{M}_\tau - \tilde{M}_\tau\right\|$ using critical values based on the distribution of $\|\mathcal{L}B - B\|$ for *any* norm $\|\cdot\|$, which are valid in the presence of heteroscedasticity of unknown form.

We propose asymptotic distribution-free tests by means of a new estimator for the integrated variance, which is of independent interest. The estimator $\tilde{\tau}$ seems applicable to testing heteroscedasticity of unknown form without prior estimation of any conditional moment using as test statistic some suitable functional of $\left\{\sqrt{n}(\tilde{\tau}(u) - \tilde{\tau}(1))\right\}_{u\in[0,1]}$. A different type of test based on a difference based estimator, and assuming independence of errors and the regressor, has been proposed by Dette and Hetzler (2009). Available functional central limit theorems for sum of functions of concomitants of order statistics, e.g. Stute (1993), and on functions of spacings, e.g. Hall (1982), could be useful for justifying the test.

Appendix. Proofs

Proof of Proposition 1. Uniformly in $u \in [0, 1]$, by definition,

$$\mathcal{L}\left(\widehat{\mathbb{M}} - \widehat{\mathbb{M}}_0\right) + \widehat{\mathbb{M}}_0 \geq \widehat{\mathbb{M}} \quad a.s.$$

Under H_0, since m is non-positive, and $\widehat{\mathbb{M}}_0$ is concave, uniformly in $u \in [0, 1]$,

$$\mathcal{L}\left(\widehat{\mathbb{M}} - \widehat{\mathbb{M}}_0\right) + \widehat{\mathbb{M}}_0 \geq \mathcal{L}\widehat{\mathbb{M}} \quad a.s.,$$

and so, uniformly in $u \in [0, 1]$,

$$\mathcal{L}\left(\widehat{\mathbb{M}} - \widehat{\mathbb{M}}_0\right) - \left(\widehat{\mathbb{M}} - \widehat{\mathbb{M}}_0\right) \geq \mathcal{L}\widehat{\mathbb{M}} - \widehat{\mathbb{M}} \quad a.s., \tag{7}$$

which implies that $\bar{\eta} \geq \hat{\eta}$ $a.s.$ Finally, (7) is satisfied with strict equality when $m(X) = 0$ $a.s.$, which implies that $\hat{\eta} = \bar{\eta}$ $a.s.$ when $m(X) = 0$ $a.s.$ ∎

Proof of Theorem 1. Henceforth, we adopt the convention that, for any sequence $\{\xi_i\}_{i\geq 1}$, $\sum_{i=2}^{j} \xi_i = 0$ for $j < 2$. Notice that, by definition of $\tilde{\tau}(u)$ and the monotonicity and continuity of $\tau(u)$,

$$\sup_{u\in[0,2/n]} |\tilde{\tau}(u) - \tau(u)| = \sup_{u\in[0,2/n]} |\tau(u)| = |\tau(2/n)| = o(1).$$

While for $u \in [2/n, 1]$,

$$\tilde{\tau}(u) = \frac{1}{2(n-1)} \sum_{i=2}^{\lfloor nu \rfloor} \left(Y_{(i:n)} - Y_{(i-1:n)}\right)^2$$

$$= \frac{1}{2(n-1)} \sum_{i=2}^{\lfloor nu \rfloor} \left(m(X_{i:n}) - m(X_{i-1:n})\right)^2 \tag{8}$$

$$+ \frac{1}{2(n-1)} \sum_{i=2}^{\lfloor nu \rfloor} \left(\varepsilon_{(i:n)} - \varepsilon_{(i-1:n)}\right)^2 \tag{9}$$

$$+ \frac{1}{(n-1)} \sum_{i=2}^{\lfloor nu \rfloor} \left(\varepsilon_{(i:n)} - \varepsilon_{(i-1:n)}\right)\left(m(X_{i:n}) - m(X_{i-1:n})\right). \tag{10}$$

We first show that (8) converges uniformly to zero *a.s.*. There exist uniformly distributed random variables on $[0, 1]$ U_i such that $X_i = F^{-1}(U_i)$ *a.s.*, and, hence, $X_{i:n} = F^{-1}(U_{i:n})$ for order statistics $U_{i:n}$. Levy (1939) showed $\max_{1\le i\le n} |U_{i:n} - U_{i-1:n}| \to 0$ *a.s.* Let $\mathcal{A}_{\max}$ be the measurable set of probability one where the latter convergence takes place. On $\mathcal{A}_{\max}$, for any $\delta > 0$ there exists an $n(\delta)$ such that for all $n \ge n(\delta)$

$$\max_{1\le i\le n} |U_{i:n} - U_{i-1:n}| \le \delta.$$

Then, since $m \circ F^{-1}$ is uniformly continuous, $\forall \epsilon > 0$, there exists a $\delta_\epsilon > 0$, such that on $\mathcal{A}_{\max}$ and for all $n \ge n(\delta_\epsilon)$,

$$\max_{1\le i\le n} \left| m \circ F^{-1}(U_{i:n}) - m \circ F^{-1}(U_{i-1:n}) \right| \le \epsilon.$$

Therefore, on $\mathcal{A}_{\max}$ and for all $n \ge n(\delta_\varepsilon)$,

$$\sup_{0\le u\le 1} \left| \frac{1}{n} \sum_{i=1}^{\lfloor nu \rfloor} \left(m(X_{i:n}) - m(X_{i-1:n})\right)^2 \right| \le \epsilon^2.$$

Since $\epsilon > 0$ was arbitrary, we conclude that (8) converges uniformly to zero *a.s.*. Next, we show that (9) converges uniformly to $\tau(u)$ in probability. Note that (9) is identical to

$$\frac{1}{2(n-1)} \sum_{i=2}^{\lfloor nu \rfloor} \left[\varepsilon_{(i:n)}^2 - \sigma^2(X_{i:n}) + \varepsilon_{(i-1:n)}^2 - \sigma^2(X_{i-1:n})\right] - \frac{1}{(n-1)} \sum_{i=2}^{\lfloor nu \rfloor} \varepsilon_{(i:n)}\varepsilon_{(i-1:n)}$$
$$+ \frac{1}{2(n-1)} \sum_{i=2}^{\lfloor nu \rfloor} \left[\sigma^2(X_{i:n}) + \sigma^2(X_{i-1:n})\right].$$

Then, it suffices to show that,

$$\left\| \frac{1}{n} \sum_{i=2}^{\lfloor n\cdot \rfloor} \varepsilon_{(i:n)}\varepsilon_{(i-1:n)} \right\|_\infty = o_{\mathbb{P}}(1), \tag{11}$$

$$\left\| \frac{1}{n} \sum_{i=1}^{\lfloor n\cdot \rfloor} \left(\varepsilon_{(i:n)}^2 - \sigma^2(X_{i:n})\right) \right\|_\infty = o_{\mathbb{P}}(1), \quad \text{and} \tag{12}$$

$$\left\| \frac{1}{n} \sum_{i=1}^{\lfloor n\cdot \rfloor} \sigma^2(X_{i:n}) - \tau(u) \right\|_\infty = o_{\mathbb{P}}(1). \tag{13}$$

Bhattacharya (1974) shows that $\left\{\varepsilon_{(i:n)}\right\}_{i=1}^n$ are conditionally independent, given $\{X_i\}_{i=1}^n$, with mean zero and conditional variance $\left\{\sigma^2(X_{i:n})\right\}_{i=1}^n$, respectively. The conditional independence implies that $S_{1k} = \sum_{i=2}^k \varepsilon_{(i:n)}\varepsilon_{(i-1:n)}$ is a martingale with respect to the filtration $\mathcal{F}_{nk} = \sigma\left(\{X_i\}_{i=1}^n, \varepsilon_{(j:n)} : 1 \le j \le k\right)$. Therefore, applying Doob's inequality, for any $\delta > 0$,

$$\mathbb{P}\left(\sup_{2\le k\le n} \left| \frac{1}{n} \sum_{i=2}^{k} \varepsilon_{(i:n)}\varepsilon_{(i-1:n)} \right| \ge \delta \right) \le \frac{1}{\delta^2 n^2} \mathbb{E}\left(S_{1n}^2\right)$$
$$= \frac{1}{\delta^2 n^2} \sum_{i=2}^{n} \mathbb{E}\left(\sigma^2(X_{i:n})\sigma^2(X_{i-1:n})\right)$$
$$= O\left(\frac{1}{n}\right).$$

In order to prove (12) we also exploit the fact that $S_{2k} = \sum_{i=1}^{k} \left(\varepsilon^2_{(i:n)} - \sigma^2(X_{i:n}) \right)$ is a martingale with respect to the filtration $\mathcal{F}_{nk}$. Hence, by Doob's inequality, for any $\delta > 0$,

$$\begin{aligned}\mathbb{P}\left(\sup_{1\le k\le n} \left| \frac{1}{n} \sum_{i=1}^{k} \left(\varepsilon^2_{(i:n)} - \sigma^2(X_{i:n}) \right) \right| \ge \delta \right) &\le \frac{1}{\delta^2 n^2} \mathbb{E}\left(S^2_{2n} \right) \\ &= \frac{1}{\delta^2 n^2} \sum_{i=1}^{n} \mathbb{E}\left[\gamma(X_{i:n}) - \sigma^4(X_{i:n}) \right] \\ &= O\left(\frac{1}{n} \right).\end{aligned}$$

Lemma 2 in Bhattacharya (1974) shows (13). Combining (11)–(13) we conclude that (9) converges uniformly to $\tau(u)$ in probability.

We show that (10) converges uniformly to zero applying Cauchy–Schwarz inequality

$$\begin{aligned}\frac{1}{n} \sum_{i=2}^{\lfloor nu \rfloor} \left(\varepsilon_{(i:n)} - \varepsilon_{(i-1:n)} \right) (m(X_{i:n}) - m(X_{i-1:n})) &\le \left[\frac{1}{n} \sum_{i=2}^{\lfloor nu \rfloor} \left(\varepsilon_{(i:n)} - \varepsilon_{(i-1:n)} \right)^2 \right]^{1/2} \left[\frac{1}{n} \sum_{i=2}^{\lfloor nu \rfloor} (m(X_{i:n}) - m(X_{i-1:n}))^2 \right]^{1/2} \\ &= o_{\mathbb{P}}(1),\end{aligned}$$

uniformly in $u \in [0, 1]$, since (8) converges uniformly to zero *a.s.* and (9) is uniformly bounded in probability. ■

Proof of Theorem 2. By Theorem 1

$$\sup_{u\in[0,1]} |\tilde{\kappa}(u) - \kappa(u)| = o_{\mathbb{P}}(1).$$

Also note that

$$\sup_{u\in[0,1]} \left| \left(\tilde{\kappa} \circ \tilde{\kappa}^{-1} \right)(u) - u \right| \le \frac{1}{n},$$

and hence,

$$\begin{aligned}\sup_{u\in[0,1]} \left| \left(\kappa \circ \tilde{\kappa}^{-1} \right)(u) - u \right| &\le \sup_{u\in[0,1]} \left| \left(\kappa \circ \tilde{\kappa}^{-1} \right)(u) - \left(\tilde{\kappa} \circ \tilde{\kappa}^{-1} \right)(u) \right| + \sup_{u\in[0,1]} \left| \left(\tilde{\kappa} \circ \tilde{\kappa}^{-1} \right)(u) - u \right| \\ &= o_{\mathbb{P}}(1).\end{aligned}$$

Bhattacharya (1974, Theorem 3) showed that $\sqrt{n}\bar{T}$ is asymptotically stochastic equicontinuous, and hence, by Remark 4.1 in Koul and Stute (1999),

$$\left\| \left(\bar{T} \circ \tilde{\kappa}^{-1} \right) - \left(\bar{T} \circ \kappa^{-1} \right) \right\|_\infty = o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}} \right). \tag{14}$$

Therefore,

$$\sup_{0\le u\le 1} \left| \left(\widetilde{\mathbb{M}}_\tau - \widetilde{\mathbb{M}}_{0\tau} \right)(u) - \frac{1}{\sqrt{\tau(1)}} \int_0^u \left(\bar{T} \circ \kappa^{-1} \right)(v) dv \right| = o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}} \right),$$

which implies, using (2), that $\sqrt{n}\left(\widetilde{\mathbb{M}}_\tau - \widetilde{\mathbb{M}}_{0\tau} \right) \to_d \mathbb{B}$ and $\sqrt{n}\bar{\eta}_\tau \to_d \zeta_0$.

Then, applying same arguments as in Proposition 1, $\tilde{\eta}_\tau \le \bar{\eta}_\tau$ *a.s.* under H_0, and $\tilde{\eta}_\tau = \bar{\eta}_\tau$ *a.s.* when $m(X) = 0$ *a.s.* This and (14) implies that, under H_0,

$$\mathbb{P}\left(\sqrt{n}\tilde{\eta}_\tau \ge c_{0\alpha} \right) \le \mathbb{P}\left(\sqrt{n}\bar{\eta}_\tau \ge c_{0\alpha} \right) \to \mathbb{P}\left(\zeta_0 \ge c_{0\alpha} \right) = \alpha \quad \text{as } n \to \infty$$

by (4) and the cmt. This proves the first part of the theorem. Under H_1, using the fact that κ^{-1} is continuous, by Theorem 1 and Glivenko–Cantelli theorem

$$\left\| \frac{\left(\hat{M} \circ \tilde{\kappa}^{-1} \right)}{\sqrt{\tilde{\tau}(1)}} - \frac{\left(M \circ \kappa^{-1} \right)}{\sqrt{\tau(1)}} \right\|_\infty = o_p(1).$$

Now notice that κ is increasing and so is κ^{-1}. Therefore, $\left(M \circ \kappa^{-1} \right)$ is non-increasing iff M is non-increasing. However, under the alternative, $\mathbb{P}(m(X) > 0) > 0$. Hence, $\left(M \circ \kappa^{-1} \right)$ is strictly increasing, and $\mathbb{M}_\tau$ convex, on some subinterval of $[0, 1]$. As a result, $\eta_\tau = \| \mathcal{L}\mathbb{M}_\tau - \mathbb{M}_\tau \| > 0$ and, hence, $\tilde{\eta}_\tau \to_p \eta_\tau > 0$ under H_1. This proves the theorem. ■

References

Andrews, D.W.K., Shi, X., 2013. Inference based on conditional moment inequalities. Econometrica 81, 609–666.
Armstrong, T., 2015. Asymptotically exact inference in conditional moment inequality models. J. Econometrics 186, 51–65.
Baraud, Y., Huet, S., Laurent, B., 2003. Adaptive tests of qualitative hypotheses. ESAIM Probab. Stat. 7, 147–159.
Baraud, Y., Huet, S., Laurent, B., 2005. Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function. Ann. Statist. 214, 214–257.
Barlow, R.E., Bartholomew, D.J., Bremmer, J.M., Brunk, H.D., 1972. Statistical Inference under Order Conditions. Wiley, New York.
Battachaya, P.K., 1984. Induced order statistics. In: Handbook of Statistics Vol. 4. pp. 383–403.
Beare, B., Moon, J.-M., 2015. Nonparametric tests of density ratio ordering. Econometric Theory 31, 471–492.
Bhattacharya, P.K., 1974. Convergence of sample paths of normalized sums of induced order statistics. Ann. Statist. 2, 1034–1039.
Blundell, R.W., Browning, M., Crawford, I.A., 2003. Nonparametric Engel curves and revealed preference. Econometrica 71, 205–240.
Bril, G., Dykstra, R., Pillers, C., Robertson, T., 1984. Algorithm AS 206: Isotonic regression in two independent variables. J. Roy. Statist. Soc. Ser. C 352–357.
Chang, M., Lee, S., Whang, Y.-J., 2015. Nonparametric tests of conditional treatment effects with an application to single-sex schooling on academic achievements. Econom. J. 18, 307–346.
Chernozhukov, V., Lee, S., Rosen, A.M., 2013. Intersection bounds: estimation and inference. Econometrica 81, 667–737.
Chetverikov, V., 2013. Adaptive test of conditional moment inequalities. Department of Economics, MIT. Unpublished Manuscript.
Chiappori, P.-A., Jullien, B., Salanié, B., Salanié, B., 2006. Asymmetric information in insurance: general testable implications. Rand J. Econ. 37, 783–798.
Cran, G.W., 1980. Algorithm AS 149: Amalgamation of means in the case of simple ordering. J. Roy. Statist. Soc. Ser. C 29, 209–211.
de Leeuw, J., Hornik, K., Mair, P., 2009. Isotone optimization in R: Pool-adjacent-violators (PAVA) and active set methods. J. Stat. Softw. 32, 1–24.
Delgado, M.A., Escanciano, J.C., 2011. Distribution-free tests of stochastic monotonicity. J. Econometrics 170, 68–75.
Delgado, M.A., Escanciano, J.C., 2013. Conditional stochastic dominance testing. J. Bus. Econom. Statist. 31 (1), 16–28.
de Paula, A., 2013. Econometric analysis of games with multiple equilibria. Annu. Rev. Econ. 5, 107–131.
Dette, H., Hetzler, B., 2009. A simple test for parametric form of the variance function in nonparametric regression. Ann. Inst. Math. 61, 861–886.
Dümbgen, L., Spokoiny, V.G., 2001. Multiscale testing of qualitative hypotheses. Ann. Statist. 29, 124–152.
Durot, C., 2003. A Kolmogorov-type test for monotonicity of regression. Statist. Probab. Lett. 63, 425–433.
Gasser, T., Sroka, L., Jennen-Steinmetz, C., 1986. Residual variance and residual pattern in nonlinear regression. Biometrika 73, 625–633.
Hall, P., 1982. Limit theorems for estimates based on inverses of spacings of order statistics. Ann. Probab. 10, 992–1003.
Hall, P., Kay, W.J., Titterington, D.M., 1990. Asymptotically optimal difference-based estimation of variance in nonparametric regression. Biometrika 77, 521–528.
Juditsky, A., Nemirovski, A., 2002. On nonparametric tests of positivity/monotonicity/convexity. Ann. Statist. 2, 85–114.
Khan, S., Tamer, E., 2009. Inference on endogenously censored regression models using conditional moment inequalities. J. Econometrics 152, 104–119.
Kim, K., 2008. Set Estimation and Inference with Models Characterized by Conditional Moment Restrictions. Preprint. University of Minnesota.
Koul, H.L., Stute, W., 1999. Nonparametric model checks for time series. Ann. Statist. 27, 204–236.
Lee, S., Song, K., Whang, Y.-J., 2013. Testing functional inequalities. J. Econometrics 172, 14–32.
Lee, S., Song, K., Whang, Y.-J., 2014. Testing for a general class of functional inequalities. Cemmap working paper CWP09/14.
Levy, P.P., 1939. Sur le division d'un segment par des points choisis au hazard. C. R. Acad. Sci. Paris 208, 147–149.
Lifshits, M.A., 1982. On the absolute continuity of distributions of functionals of random processes. Theory Probab. Appl. 27, 600–607.
Pakes, A., 2010. Alternative models for moment inequalities. Econometrica 78, 1783–1822.
Paley, R.E., Wiener, N., 1934. Fourier Transforms in the Complex Domain, Vol. VIII. Amer. Math. Soc., New York.
Rice, J., 1984. Bandwidth choice for nonparametric regression. Ann. Statist. 4, 1215–1230.
Stute, W., 1993. U-functions of concomitants of order statistics. Probab. Math. Statist. 14, 143–155.
von Neumann, J., Kent, R.H., Bellinson, H.R., Hart, B.I., 1941. The mean square successive difference. Ann. Math. Statist. 12, 153–162.